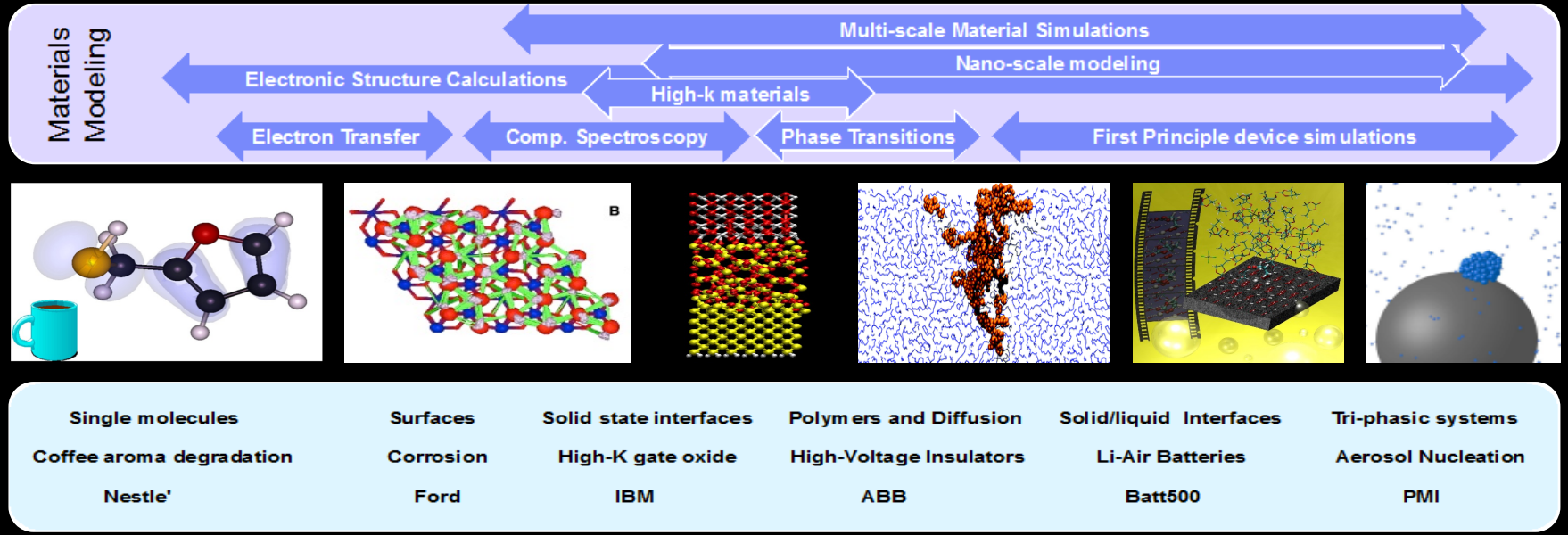
Machine Learning and Artificial Intelligence: applications to Materials Science for customized solutions

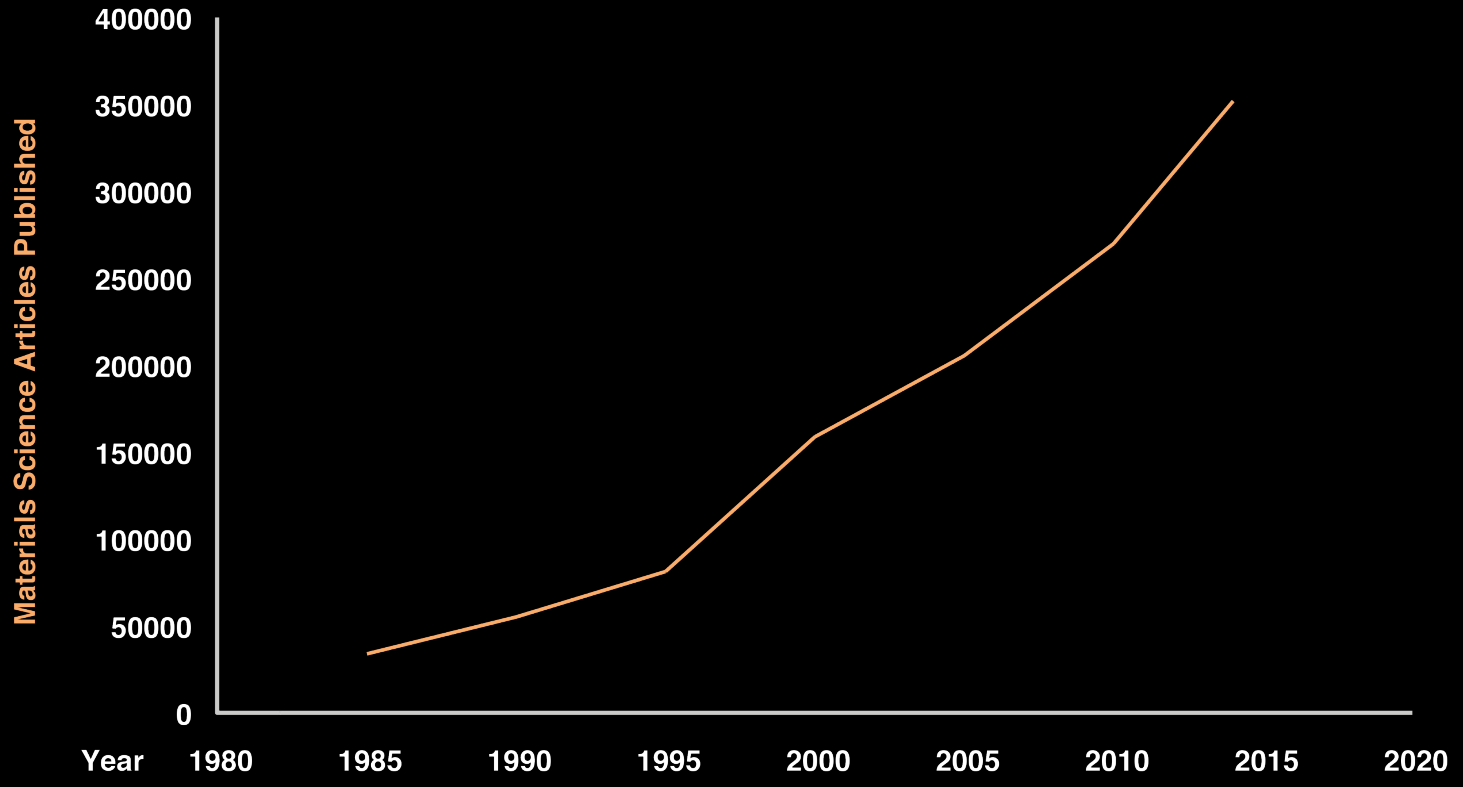
July 29, 2019
Materials Science in Space Workshop (ISSR&D)

Teodoro LAINO, PhD
Principal RSM, IBM Research - Zurich

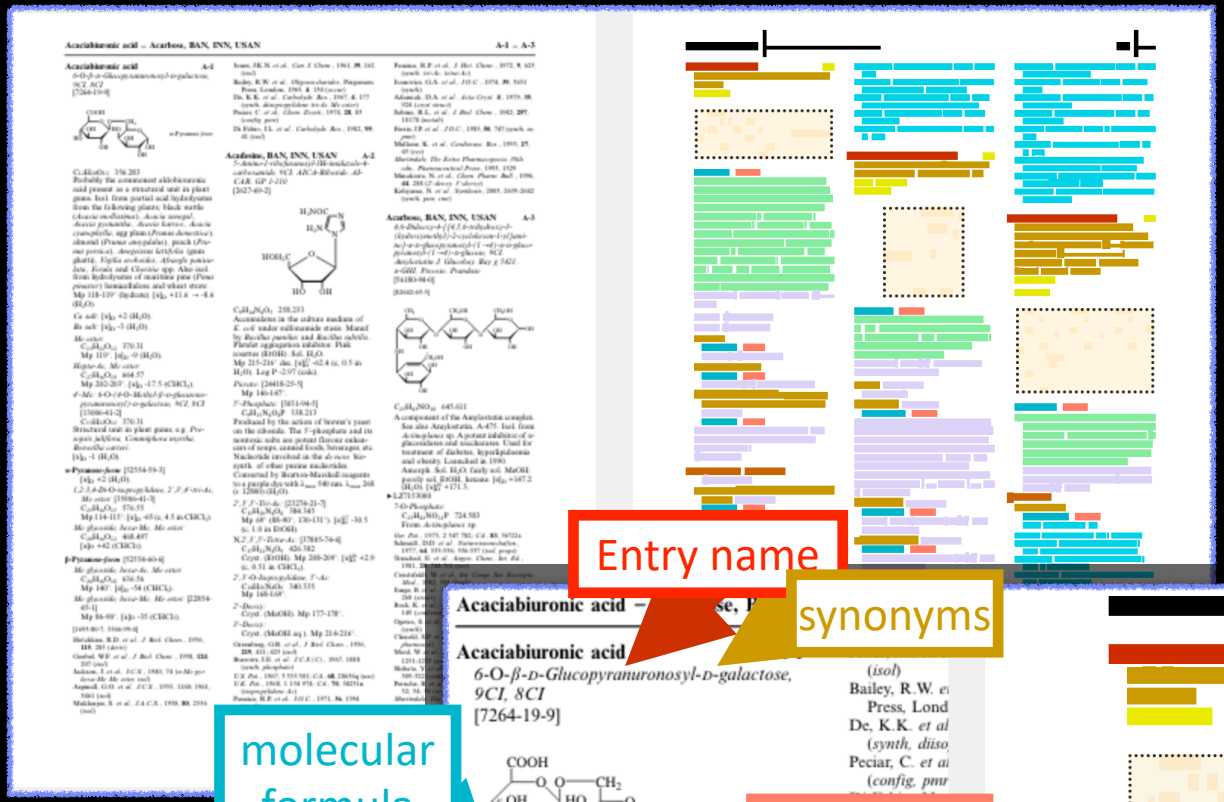
HPC Enabled Simulations of Realistic Models



Unstructured Data Deluge in Peer Review Publications



Ingesting PDF



PDF Parser

Handbook of Carbohydrates

Entry name

synonyms

molecular formula

molecular weight

Acaciabiuronic acid

6-O-β-D-Glucopyranuronosyl-D-galactose, 9CI, 8CI

[7264-19-9]

$C_{12}H_{20}O_{12}$ 356.283

Probably the commonest aldobiouronic acid present as a structural unit in plant gums. Isol. from partial acid hydrolysates from the following plants; black wattle (*Acacia mollissima*), *Acacia senegal*, *Acacia pycnantha*, *Acacia karroo*, *Acacia cyanophylla*, egg plum (*Prunus domestica*),

Subtitle-level-

Text

None

chem-formula

properties

description

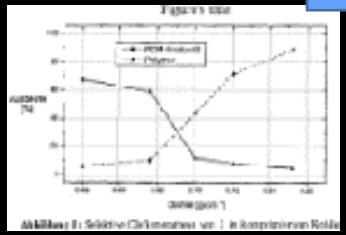
prop-Mp

prop-mol_wt

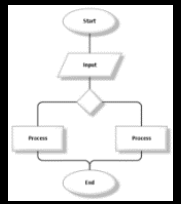
Document Understanding Workflow

PDF Parser

Image classification

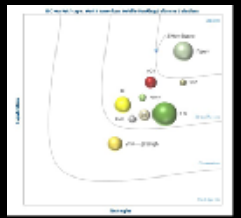


line and scatter plots



flow charts

molecular pathways



bubble plots

Major New York Bridges

Bridge	Designer	Length
Brooklyn	J. A. Roebling	1595
Manhattan	G. Lindenthal	1470
Williamsburg	L. L. Buck	1600
Queensborough	Palmer & Hayward	1182
Triborough	O. H. Ammann	1380 383
Bronx Whitestone	O. H. Ammann	2300
Throgs Neck	O. H. Ammann	1800
George Washington	O. H. Ammann	3500

scanned tables

Tables understanding

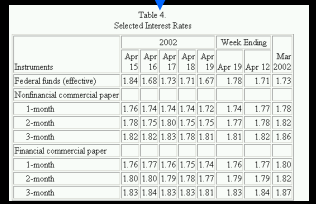
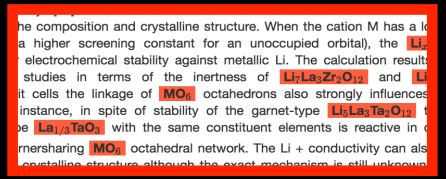


Table 4.
Selected Interest Rates

Instruments	2002 Apr 15	Apr 16	Apr 17	Apr 18	Apr 19	Week Ending Apr 19	Apr 12	Mar 2002
Federal funds (effective)	1.84	1.68	1.73	1.71	1.67	1.78	1.71	1.73
Nonfinancial commercial paper								
1-month	1.76	1.74	1.74	1.74	1.72	1.74	1.77	1.78
2-month	1.78	1.75	1.80	1.75	1.75	1.77	1.78	1.82
3-month	1.82	1.82	1.83	1.78	1.81	1.81	1.82	1.86
Financial commercial paper								
1-month	1.76	1.77	1.76	1.75	1.74	1.76	1.77	1.80
2-month	1.80	1.80	1.79	1.78	1.77	1.79	1.79	1.82
3-month	1.83	1.84	1.83	1.83	1.81	1.83	1.84	1.87

Text analytics NLP



he composition and crystalline structure. When the cation M has a l
a higher screening constant for an unoccupied orbital), the Li
electrochemical stability against metallic Li. The calculation results
studies in terms of the inertness of $Li_7La_3Zr_2O_{12}$ and Li
it cells the linkage of MO_6 octahedrons also strongly influences
instance, in spite of stability of the garnet-type $Li_5La_3Ta_2O_{12}$
be $La_{1/3}TaO_3$ with the same constituent elements is reactive in
rnersharing MO_6 octahedral network. The Li + conductivity can als
crystalline structure although the exact mechanism is still unknown

Semantic analysis: unsupervised generic extraction

text → "At ambient temperatures, **tungsten** is **not attacked** by hydrochloric, sulfuric nitric, and hydrofluoric acids at any concentration, or by **aqua regia** "

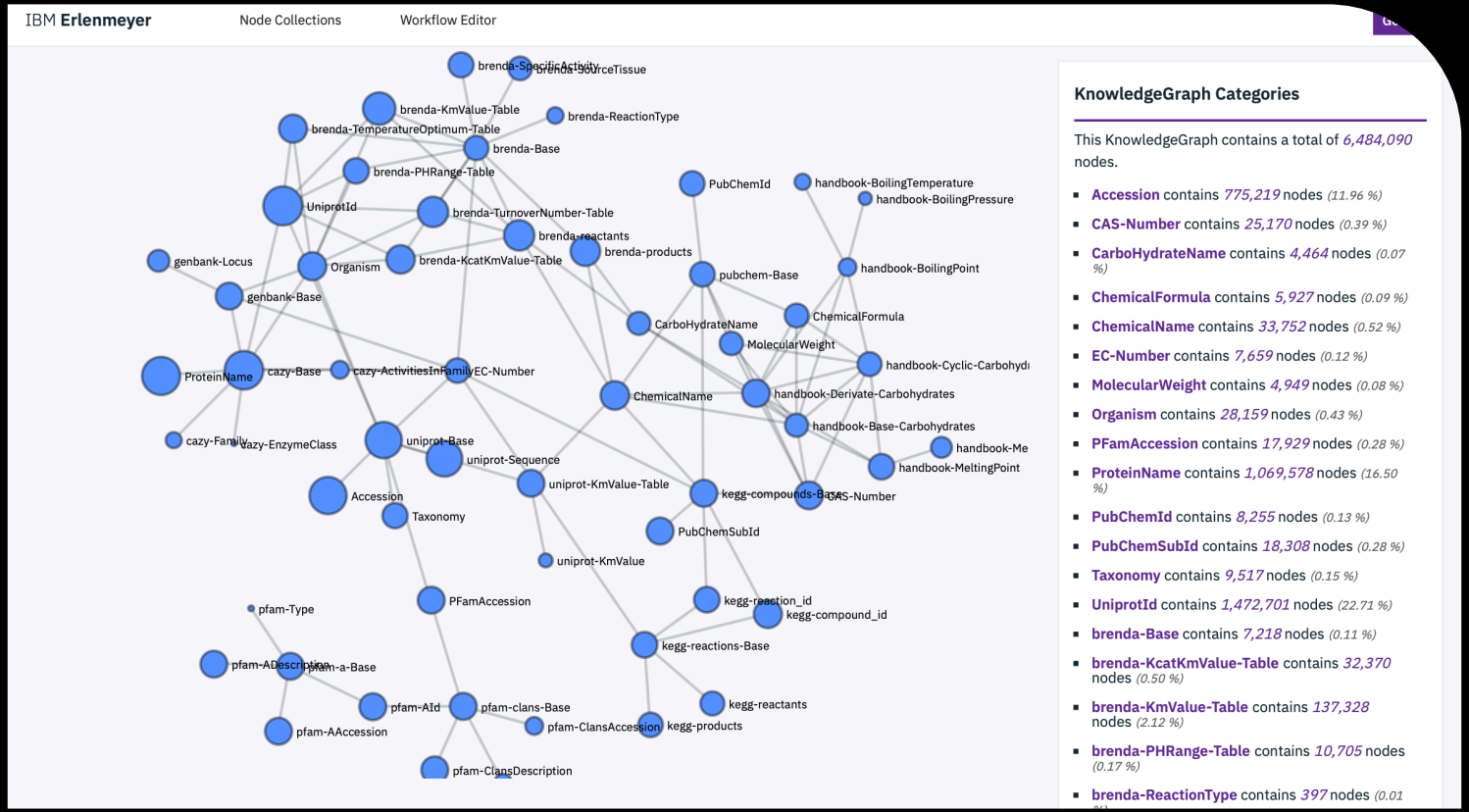
REL GENERIC TRIPLETS

tungsten // **to_be__not__attack ambient__temperature** // hydrochloric__acid
tungsten // **to_be__not__attack ambient__temperature** // sulfuric__acid
tungsten // **to_be__not__attack ambient__temperature** // nitric__acid
tungsten // **to_be__not__attack ambient__temperature** // hydrofluoric__acid
tungsten // **to_be__not__attack ambient__temperature** // aqua_regia

ADDITIONAL TRIPLETS

hydrochloric // HasType // chemical_substance,_chemical_compound
nitric // HasType // chemical_substance,_chemical_compound
hydrofluoric // HasType // chemical_substance,_chemical_compound

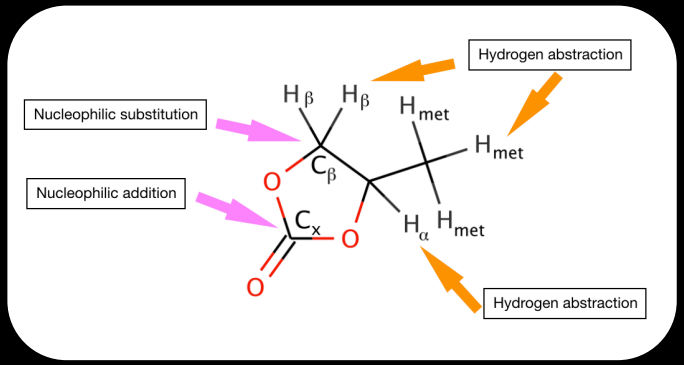
Domain Knowledge Graph



A real case: base-line

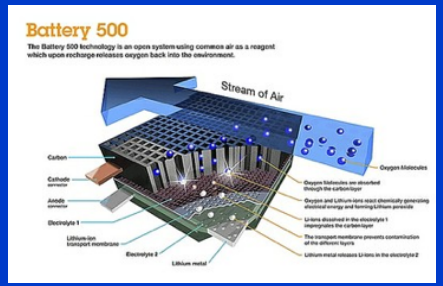
The Propylene Carbonate (PC) may degrade in different possible ways.

In 2015, we addressed this problem by doing extensive and CPU-intensive DFT calculations.



Battery 500

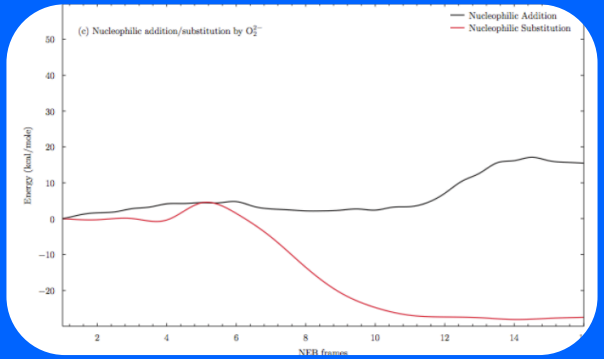
Organic liquid electrolyte chemical stability issue



Approximately 2 months to complete the analysis of the possible degradation pathways.

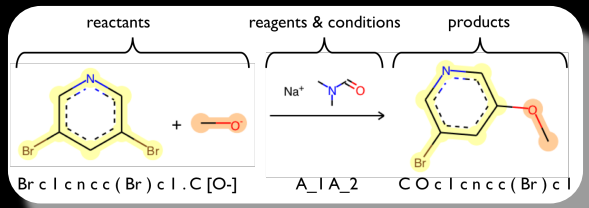
~ 5 Mio CPU-hours (BGQ) to generate reactive profiles (NEB) + 100 man-hours for analysis and simulation assistance

An example of an NEB outcome
From reactants to products **(PC degraded by peroxide)**

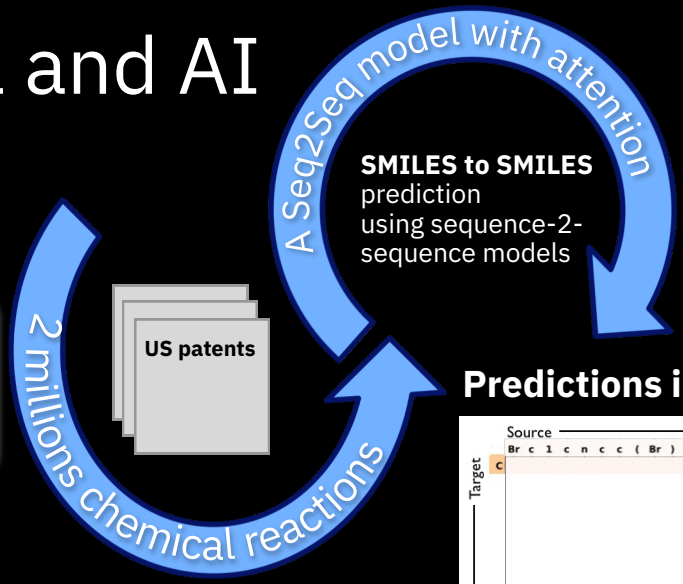


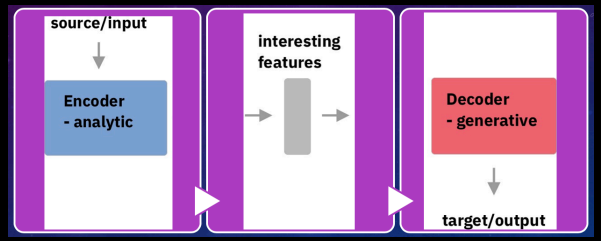
A real case: Data and AI

Forward Chemical reaction prediction

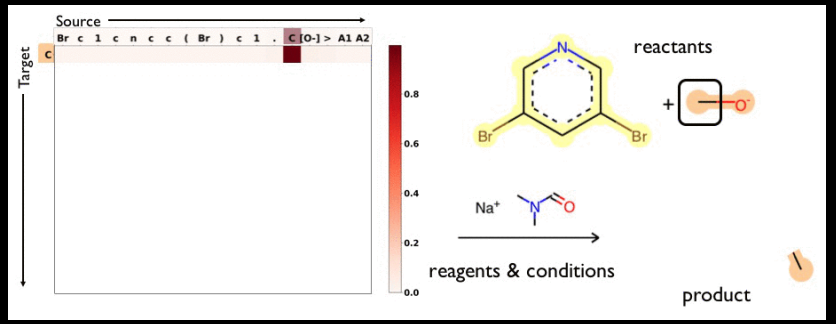


SMILES to SMILES





Predictions in short (how a Se2Seq works)



Schwaller et. al. Chem. Sci., 2018, 9, 6091-6098

Freely available for everyone:
http://www.research.ibm.com/ai4chemistry

Released on August 19th at ACS Boston. In 10 months (Jun. 2019):

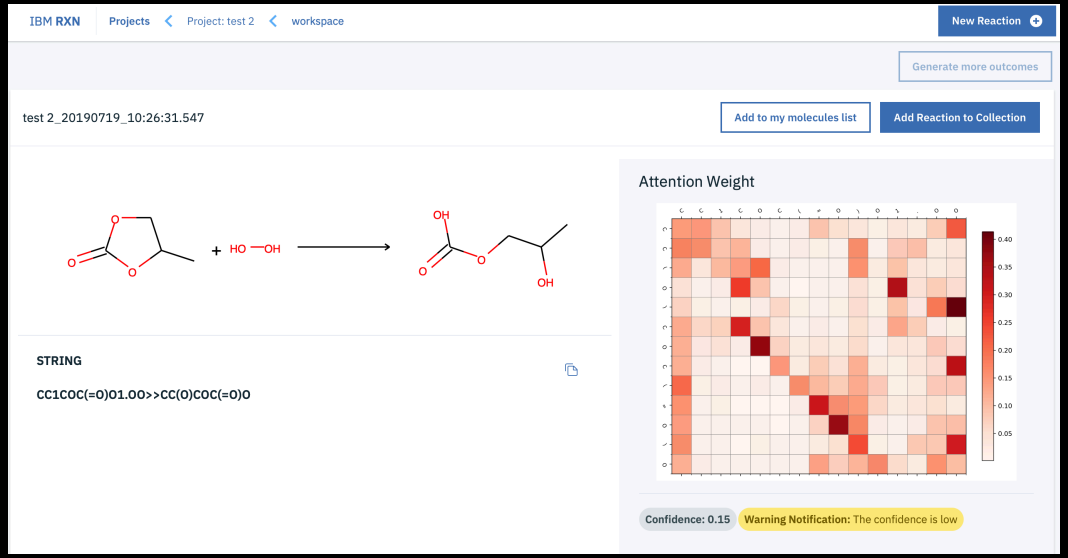
- 6300 users
- 48000 chemical reactions predictions

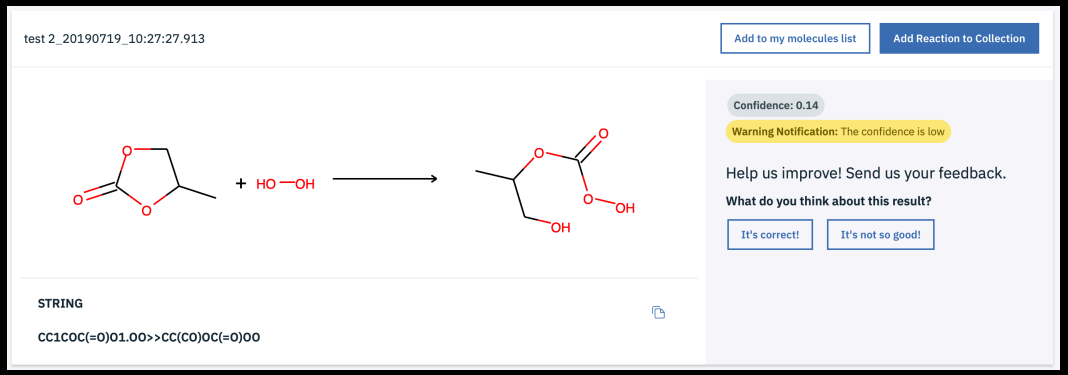
Performance

USPTO*		S2S [13]	WLDN [28]	ELECTRO [30]	GTPN [31]	WLDN5 [29]	our work
_MIT	separated	80.3	79.6		82.4	85.6	90.4
_MIT	**mixed**		74				**88.6**

Currently, best method in FORWARD chemical reaction prediction

A real case: AI benefits





< 6 seconds prediction time for possible degradation products

Correct prediction order (Substitution more favorable than Addition)

Screening 30 different electrolytes using DFT took approximately 6 months (or **15 Mio seconds**). Using the AI model (RXN) for an initial screening this would have taken **300 seconds**. Accelerated by **50,000**.

Thank you.

Questions ?